BBC RD 1996/4

Research and Development Report

ISO/IEC MPEG-2 AUDIO: Bit-rate-reduced coding for two-channel and multichannel sound

G. Stoll, Dipl.-Ing., (Institute für Rundfunktechnik, Munich, Germany)
and N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E., AES Fellow

Research & Development Department
Policy & Planning Directorate
THE BRITISH BROADCASTING CORPORATION

*ISO/IEC MPEG-2 AUDIO: Bit-rate-reduced coding for two-channel and multichannel sound

G. Stoll, Dipl.-Ing., (Institut für Rundfunktechnik, Munich, Germany)
***and* N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E., AES Fellow**

Summary

Since 1988 the ISO/IEC Moving Picture Experts' Group (MPEG) has been developing generic coding standards for video and audio, mainly for broadcast and multimedia applications. The MPEG audio standard (ISO/IEC 11172-3) was finalised in November 1992. It follows a three-layer structure in order to fulfil the requirements of various applications. Good audio quality at a bit rate of about 130 kbit/s per monophonic channel is achieved.

The first objective of MPEG-2 audio was the extension from two to five channels, based on standards and recommendations from international organisations such as ITU-R, SMPTE andEBU. This was achieved in November 1994 with the approval of the ISO/IEC document13818-3, sometimes termed 'MPEG-2 Audio'. This standard provides high qualitycoding of 5 full-bandwidth audio channels plus an optional low-frequency effects (LFE or 'sub-woofer') channel, together with backwards compatibility to MPEG-1. Backwards compatibility is the key to ensuring that existing 2-channel decoders will still be able to decode compatible stereo information from multichannel signals. For audio reproduction of surround sound, the loudspeaker positions left, centre, right, with left and right surround are used (according to the 3/2-standard). The envisaged applications are digital television systems such as dTTb, HDTV, HD-SAT, ADDT, as well as digital audio broadcasting (DAB) and digital storage media. The second objective was the extension of MPEG-1 audio to lower sampling frequencies to improve the audio quality at bit rates less than 64 kbit/s per channel, in particular for commentary applications.

* The core material of this Report is an expansion of the copyright paper by G.Stoll (IRT), written for the Audio Engineering Society (AES) Special Publication:- *Collected Papers on Digital Audio Bit-rate Reduction*, and is published with permission by the AES and the author.

Issued under the Authority of

Research & Development Department
Policy & Planning Directorate
BRITISH BROADCASTING CORPORATION

General Manager
Research & Development Department

ISO/IEC MPEG-2 AUDIO:
Bit-rate-reduced coding for two-channel and multichannel sound

G. Stoll, Dipl.-Ing., (Institut für Rundfunktechnik, Munich, Germany)
and N.H.C. Gilchrist, B.Sc., C.Eng., M.I.E.E., AES Fellow

ISO-MPEG-2 AUDIO: A generic standard for the coding of two-channel and multichannel sound

G. Stoll, Dipl.-Ing., (Institut für Rundfunktechnik, Munich, Germany)
and **N.H.C. Gilchrist,** B.Sc., C.Eng., M.I.E.E.

1. INTRODUCTION

Digital audio was introduced to the consumer in the early 1980s with the Compact Disc (CD). The 16-bit PCM format of the CD is an accepted audio reproduction standard, although the bit-rate of about 706 kbit/s per mono channel is rather high. Lower bit-rates are essential if there is only a limited capacity available for the storage or transmission of audio signals. Typical application areas for low bit-rate coded digital audio are:

- Programme distribution and exchange
- Digital audio broadcasting (DAB)
- Digital storage (e.g. archiving, studio recording and consumer electronics)
- Interpersonal communications such as video-conferencing and multimedia applications
- Enhanced-quality TV systems.

New coding techniques for high quality audio signals use the properties of human sound perception by exploiting the spectral and temporal masking effects of the ear. The objective is for the quality of the reproduced sound to be as good as that obtained with 16-bit PCM at a sampling frequency of 44.1 or 48 kHz, whilst using a minimal bit-rate for the coded signal Such a source coding system was recently standardised by ISO/IEC.[1] It allows the bit-rate of a 16-bit digital audio signal sampled at 48 kHz to be reduced from 768 kbit/s (about 706 kbit/s for a sampling frequency of 44.1 kHz) to approximately 128 kbit/s per mono channel, while preserving the subjective quality of the original signal. This reduction in bit-rate coding is possible because the coding adapts the quantising noise to the masking characteristics of the original audio signal, and only those details of the signal which will be perceived by the listener are transmitted.

Fig. 1 shows the wide range of different bit-rates needed for a number of techniques for mono and stereo audio digital coding.

When considering digital audio for television, it is clearly necessary to determine what sound system should be adopted and what level of service will be provided to the consumer. Should a completely new system be provided, or should an existing proposal be adapted for digital television? Are consumers to be offered just mono and stereo, or should multichannel sound (surround sound and multilingual services) be made available? In either case, one must take account of the consequent demands on the programme channel (i.e. the bit-rate required).

For digital television, the choice of the audio coding system will depend upon such things as:

- The available bit-rate
- Compatibility with other services and hardware (both within and between countries)

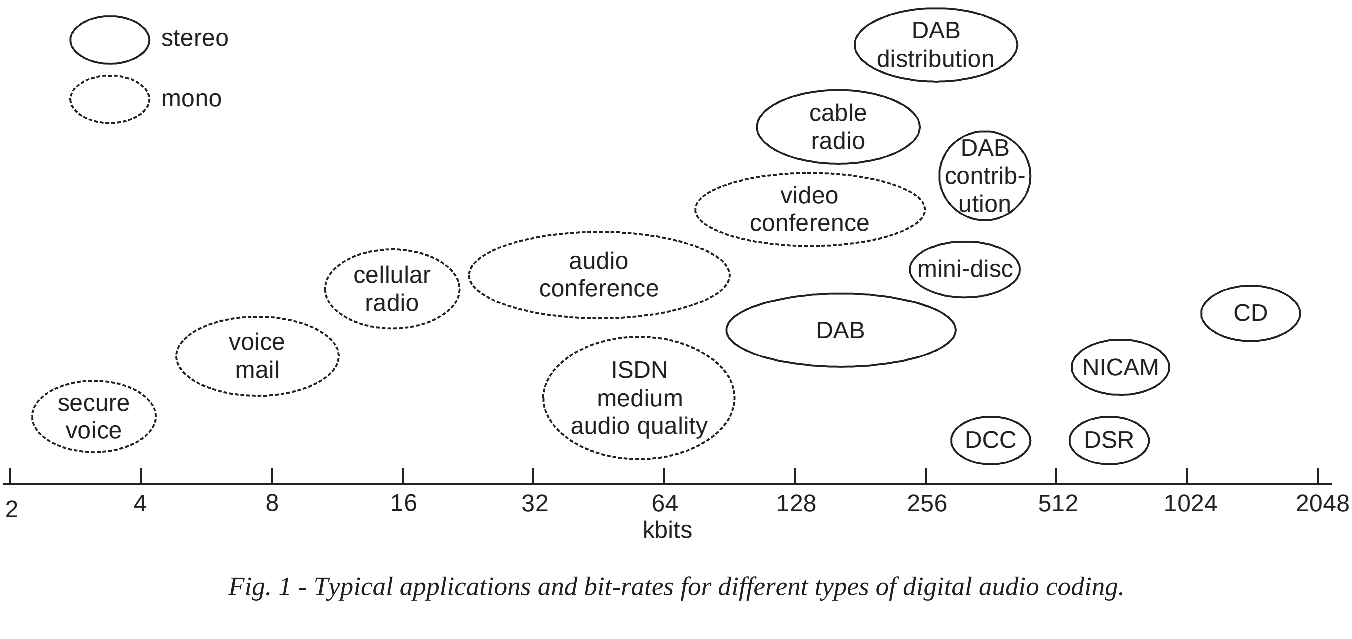

Fig. 1 - Typical applications and bit-rates for different types of digital audio coding.

- The need for reconfiguration of the transmitted signals
- The possible need to provide performance margins for multiple coding and decoding of the signal by tandem-connected codecs
- The services to be offered to the consumer.

Two further MPEG-2 topics – ‘Video coding’[2] and ‘Overview of the systems layer’[3] – are available as companion R&D Reports.

2. DIGITAL TELEVISION SOUND SERVICES

Members of specialist international groups* have spent much time evaluating the service options that can be considered for multichannel sound. Although the discussions have centred mainly upon the options for an HDTV service, they have considered other television service situations and audio-only applications. The two main options considered have been the ones for surround sound and for multilingual services.

The surround sound option has been considered as an important one for several reasons. Firstly, surround sound is already available within the cinema industry, and to some degree also, within the broadcasting industry – where a stereo transmission format is available and some source material is already in analogue Dolby Stereo** form (e.g. film sound tracks and some sports events). Secondly, if surround sound is to be introduced into a digital broadcasting environment, something operationally more suitable than the analogue method of combining channels needs to be developed. Thirdly, domestic equipment manufacturers participating within specialist collaborative projects*** have expressed a strong belief that significant improvements in the sound system will be essential in the marketing of new, more expensive television systems.

Subjective tests, carried out specifically to quantify the subjective benefit of the different forms of reproduction,[4,5] have shown that going from mono (1/0) to stereo (2/0) was equivalent to one grade improvement on the ITU-R 5-point quality grading scale.[6] Going from stereo (2/0) to three channel (3/0) (i.e. with a centre channel) was equivalent to an additional one grade of improvement, and from three channel (3/0) to surround sound (3/2) was equivalent to an additional grade improvement.

* Eureka 95; European Multichannel Experts Group (EUMEG); ‘Grand Alliance’ in the USA; MPEG; Task Group 10-1 of the international Telecommunication Union’s Radiocommunication Sector (ITU-R).

** ‘Dolby Stereo’ is a Trademark of Dolby Laboratories Licensing Corporation.

*** HD-SAT (High Definition Television – Satellite); dTTb (Digital Terrestrial Television Broadcasting; HDTV-T (High Definition Television –Terrestrial); Eureka 1187 ADTT (Advanced Digital Television Technologies); USSB (United States Satellite Broadcasting); DSS (Digital Satellite System); DirecTV.

3. MAJOR DIGITAL AUDIO CODING SYSTEMS

3.1 Coding used for broadcasting and storage applications

The main work in development and standardisation, including extensive subjective evaluation, has taken place on low bit-rate audio codecs within the various international groups (MPEG, ITU-R, EBU, Grand Alliance etc.). This work has led to conclusions and system proposals that are applicable to the different constraints within which each group has been working.

High quality audio coding schemes already have, and will continue to have, many applications in the areas of recording, computer multimedia, telecommunication, radio and television broadcasting, cable and film. Table I provides an overview of the major digital two-channel and multichannel audio coding systems already in use or proposed for various applications. The systems are presented in alphabetical order.

Three of the coding schemes mentioned are used already in the area of home recording.† In the telecommunications area, ITU-T standards, such as G711 and G722, are being replaced by MPEG-1 audio Layer II and Layer III coding. Computer multimedia uses mainly MPEG-1 audio Layer II. The older digital radio satellite services, such as the German DSR-system, use simple block-companding techniques,[8] but more recent satellite and terrestrial digital radio services within Europe use Layer II. The AT&T PAC system[9] is proposed for one of the USA digital systems.

In television broadcasting, NICAM[10] has been introduced in some European countries to provide digital stereo sound. For future television systems, the European broadcasters show a clear preference for the MPEG-1[1] and MPEG-2[11] interrelated standards, whereas the North American broadcasters are currently preferring two differing standards, namely MPEG-1 (DirecTV) and Dolby AC-3[12] (Grand Alliance).

3.2 Coding formats used in the film industry

A dilemma for the film industry is currently posed by the existence of four different and completely incompatible formats for digital audio. Two of the systems,

† ATRAC system for Sony’s MidiDisc,[7] MPEG-1 audio coding[1] (Layer I is used for Philip’s DCC, and Layer video-CD).

Sony's SDDS (Sony Dynamic Digital Sound) and DTS from Digital Theatre Systems (a six channel surround system with the audio stored on a CD-ROM), are associated with major film studios. DTS is associated with Universal-MCA and SDDS with Sony Pictures (Columbia and TriStar). The third, Dolby AC-3, also known as SR-D or Dolby Stereo Digital, is independent. The fourth system for digital film sound, used mainly in France and based on MPEG-1 Audio Layer II, is Cinema DSP. If a single standard does not emerge, this situation will pose rather a predicament for studios and cinemas alike.

First of all, most of the cinema chains want to have digital multichannel sound for their auditoria, but they will not want to purchase three or four systems for each of their premises. MGM/United Artists are directing their attention towards DTS. New Line Cinema have expressed a qualified endorsement of DTS, because theatre owners are more likely to favour DTS because of the cost, compared to Dolby AC-3. Sony Pictures has stated that all films where Sony has control over the distribution will be in SDDS. However, even Castle Rock, which uses Sony for distribution, is releasing some pictures in DTS, and Paramount recently announced an agreement for five films in DTS. Although DTS may seem to be the front-runner, it is not a clear victor. This is a situation which will probably exist for some time yet.

The fact that the film industry has not selected one digital audio system poses a dilemma for broadcasters, because transcoding between the film sound format and the broadcast format cannot be avoided, whichever digital audio system is chosen for broadcasting.

4. ISO/MPEG AUDIO: GENERIC CODING OF STEREO AND MULTICHANNEL SOUND

From 1988 to 1992 the International Organisation for Standardisation (ISO) has been developing and preparing a standard on information technology – coding of moving pictures and associated audio for digital storage media up to about 1.5 Mbit/s.[1] The 'Audio Subgroup' of MPEG had the responsibility for developing a standard for generic coding of PCM audio signals with sampling rates of 32, 44.1 and 48 kHz at bit-rates in a range from 32 kbit/s to 192 kbit/s per mono channel and 64 to 384 kbit/s for stereo signals.

Two mechanisms can be used to reduce the bit-rate of audio signals. One mechanism removes the redundant information from the audio signal. The other removes the irrelevancy of the audio signal by taking advantage of psychoacoustic phenomena, like spectral and temporal masking. Only with both of these techniques, exploiting redundancy and the masking effects of the

Table 1:
Major digital audio coding systems for single channel, two-channel and multichannel applications.

Coding scheme	Type	Mono	Stereo	Multichannel	Half sampling rate mono/stereo
APT[13]	apt x-100		•		
AT&T	PAC		•		
AT&T	MPAC			•	
Dolby	AC-2		•		
Dolby	AC-3	•	•	•	
DSR	Block companding	•	•		
MPEG-1	Layer I	•	•		
MPEG-1	Layer II	•	•		
MPEG-1	Layer III	•	•		
MPEG-2	Layer I	•	•	•	•
MPEG-2	Layer II	•	•	•	•
MPEG-2	Layer III	•	•	•	•
NICAM	Block companding	•	•		
PCM 16 BIT	Linear representation	•	•		
SONY	ATRAC		•		

human ear, can a significant reduction of the bit-rate, down to about 200 kbit/s per stereophonic signal, be obtained.

Different layers of the coding system with differing degrees of encoder and decoder complexity and performance are described in the audio part of ISO Standard 11172 (MPEG-1).[1] The idea behind the three-layer concept was to have a universal coding scheme for many applications with totally different requirements. These would include consumer recording, professional recording, the combined recording and processing of audio and video, telecommunications and broadcasting. A comprehensive description of the whole MPEG-1 audio coding standard with, a detailed explanation of its layer concept, can be found elsewhere.[14]

4.1 MPEG-2 audio: generic multichannel audio coding

The first objective of MPEG-Audio phase 2, sometimes known as MPEG-2 Audio, was the development of a standard for the low bit-rate coding of multichannel audio; using perceptual coding methods which can be used to transfer high quality digital surround sound and/or multilingual audio information on channels with limited capacity.[15] The MPEG-2 audio standard[11] was approved by the MPEG committee in November 1994. It takes account of standards and recommendations from international organisations such as ITU-R, SMPTE (the Society of Motion Picture and Television Engineers) and the EBU. One important requirement is backwards compatibility to mono, stereo or dual-channel audio programmes coded in accordance with ISO/IEC 11172-3.[1] To fulfil this requirement, the coded signal must be such that an ISO/IEC MPEG-1 audio decoder is able to correctly decode the basic stereo information from the multichannel programme. The basic stereo information needs to be kept in the frontal left- and right-channel components, which constitute an appropriate downmix of the audio information in all channels.

The backwards compatibility requirement arises because many integrated decoding chips for ISO/MPEG audio and video are under development. The audio decoders will handle only two audio channels, however. With backwards compatibility, a two-channel decoder will be able to deliver a basic stereo signal from the multichannel audio bitstream.

The main techniques for reducing the bit-rate of MPEG-2 multichannel coding, while maintaining the subjective quality of the input audio signals, are: sub-band filtering, perceptual modelling of the ear, the sharing of bits between channels from a common pool, joint stereo coding, common masking thresholds and the introduction of dynamic crosstalk between channels.

The second objective of MPEG-2 Audio was the extension of MPEG-1 Audio to lower sampling frequencies to improve the audio quality for mono and conventional stereo signals for bit-rates at or below 64 kbit/s per channel; in particular, for commentary applications. This goal has been achieved by reducing the sampling frequency to 16, 22.05 or 24 kHz, with a consequent reduction in the audio bandwidth to 7.5, 10.5 or 11.5 kHz. Compared with MPEG-1, the only differences in the encoder and decoder, other than the change in the clock rate, are changes in the encoder and decoder tables of bit-rates and bit allocation. The encoding and decoding principles of MPEG-1 Audio Layers I, II and III are fully maintained. Table II shows the main areas of MPEG-2 audio work.

4.1.1 Extension of MPEG audio to multichannel coding

At present, multichannel audio is known primarily from the cinema. However, even in consumer applications, multichannel audio has been available for the last few years (e.g. Dolby Surround with domestic television and videocassette recorders). With the introduction of Advanced or High Definition Television (ADTV, HDTV) having improved resolution and increased picture size, to give an improved visual perspective more like a cinema, improved realism from the audio is appropriate. A way to achieve this is to use more than two audio channels. Recently, ITU-R, SMPTE and EBU have started to standardise the listening arrangement for loudspeaker reproduction of multichannel audio. An advantage of this system is the relatively large area over which satisfactory listening can be experienced. There is, however, the disadvantage that a relatively high bit-rate is needed for the digital audio signals to be transmitted or recorded. With the application of the coding system described in this paper, economical digital storage or transmission of the multichannel audio is possible. In addition to ADTV and HDTV, a number of multimedia applications will adopt multichannel audio if good performance can be obtained economically at low data rates.

4.1.2 Characteristics of the MPEG-2 audio coding system

A generic digital multichannel sound system applicable to television and sound broadcasting and storage, as well as other non-broadcasting media, should meet several basic requirements and provide a number of technical/operational features. In addition to two-channel compatibility and interoperability between different media, and downwards compatibility with sound formats consisting of a smaller number of audio channels, other aspects also need consideration. Multilingual services, ‘clean’ dialogue and dynamic range compression are important, in order to serve the widest possible range of applications. It is important,

Table II:
Main work areas of MPEG-2 audio.

1	**Backwards compatible multichannel sound**	
	♦ Extension of the ISO/MPEG-1 standard up to five audio channels, plus a low-frequency enhancement channel. ♦ *Backwards compatibility:* A current MPEG-1 stereo decoder will reproduce a compatible stereo signal, i.e. a down-mix of all five channels, when supplied with an MPEG-2 bit stream. Both the programme provider and the the consumer can switch from two-channel to multichannel at any time.	♦ Up to seven multilingual channels, either with the same or with half the sampling frequency of the main audio programme. ♦ The MPEG-2 Audio standard was finalised in November 1994.
2	**Extension to lower sampling frequencies**	
	♦ Three additional sampling frequencies: 16, 22.05 and 24 kHz. ♦ Reduced bandwidth (up to 11.5 kHz) but improved coding gain, with better adaptation to the masking threshold, giving much better audio quality at bit-rates below 64 kbit/s per channel. Especially suitable for commentary applications.	♦ Excellent commentary quality at 48 kbit/s, together with 16 kbit/s for ancillary data in an ISDN B-channel of 64 kbit/s. ♦ Only minor changes are needed to the ISO/MPEG standard (two tables in the decoder). The current encoder and decoder hardware and software can easily be updated to support all six sampling frequencies
3	**Non-backwards compatible multichannel sound**	
	♦ Requirements defined during March-July 1994. ♦ Collaborative effort will lead to the combination of the best algorithmic elements of all submissions.	♦ An addendum to the ISO/MPEG-2 standard is due to be finalised in July 1997.

as well, to obtain a level of audio quality close to that of the original signal; typically, a linearly-coded PCM audio signal with a resolution of at least 16 bits, without requiring an unreasonably complex decoder.

MPEG-2 audio provides for a wide range of bit-rates from 32 kbit/s up to 1066 kbit/s. If the audio is subjected to only one coding/decoding process, quite a low bit-rate may give acceptable results. It is of considerable importance to digital audio broadcasting (DAB), for example, to be able to use really low bit-rates, because of the relatively low transmission capacity. Higher rates, up to about 180 kbit/s per channel may be necessary if there are a number of coding/decoding processes or there is the need for some post-processing (e.g. a re-mix).

4.1.3 3/2-stereo presentation performance

As regards stereophonic presentation, specialist groups of the ITU-R, SMPTE and EBU recommend a 5-channel system as the reference surround sound format.[16] This arrangement has a centre channel C and two surround channels Ls, Rs, in addition to the front left and right stereo channels L and R. It is referred to as ‘3/2-stereo’ (3 front/2 surround channels) and requires the handling of five channels in all situations (the studio, storage media, contribution, distribution, emission links, and in the home). Fig. 2 shows the positions of the loudspeakers reproducing the 5-channel signal in a reference listening arrangement.

For sound that accompanies pictures, the three frontal channels ensure directional stability and clarity of the picture-related frontal images, and accords to common practice in the cinema. In particular, the centre channel is of high importance for certain situations (e.g. where

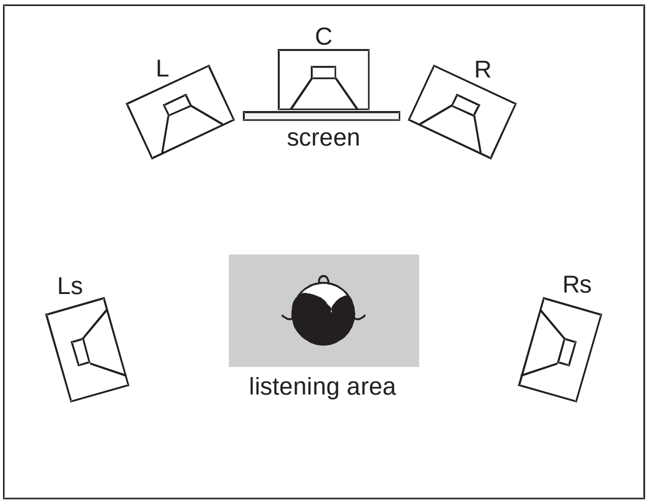

Fig. 2 - 3/2 reference loudspeaker arrangement.

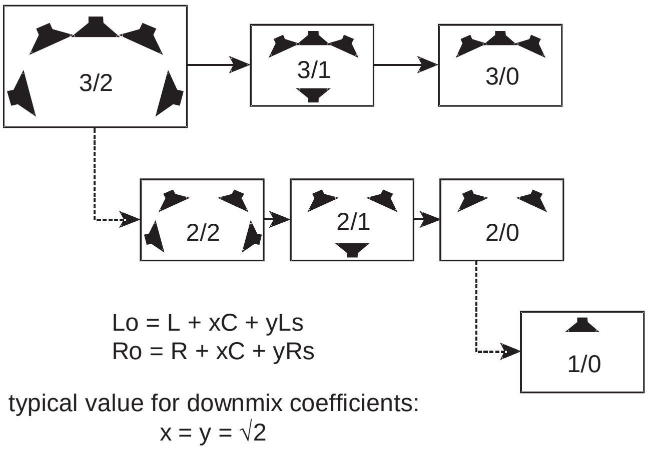

Fig. 3 - Downmix from 3/2 multichannel to 1/0 mono with MPEG-2 audio coding.

a dialogue requires stable localisation in the middle of the frontal area). Additionally, such a centre channel enlarges the area of the listening zone. For audio-only applications three frontal channels have been found to give a worthwhile improvement over two-channel stereophony.[4,5] The addition of one pair of surround channels to the three front channels improves realism.

Optionally, there may be an even number of more than two rear/side loudspeakers which may provide a larger optimum listening area. Since the locations of the side/rear surround loudspeakers are largely non-critical with respect to both direction and distance, they should be accommodated readily in an existing living-room environment.

4.1.4 Downward compatibility

A hierarchy of sound formats providing a lower number of channels and reduced stereophonic presentation performance (down to 2/0-stereo or even mono), and a corresponding set of downward mixing equations can be recommended,[17] to provide downward compatibility, as shown in Fig. 3. Useful alternative lower level sound formats are 3/0, 2/2, 2/0, 1/0. These may be used in circumstances where economic or channel capacity constraints apply in the transmission link, or where a lower number of reproduction channels is required.

4.1.5 Backward/Forward compatibility with MPEG-1

For several applications, it is the intention to improve the existing 2/0-stereo sound system progressively by transmitting additional sound channels (centre, surround) without making use of a simulcast operation. The multichannel sound decoder has to be backwards/forwards compatible with the existing sound format.

Backwards compatibility means that an existing two-channel (low price) decoder should properly decode the basic 2/0-stereo information from the multichannel bitstream (see section 4.1.6). This implies the provision of compatibility matrices using appropriate downmix coefficients, as shown in Fig. 4.

Forwards compatibility means that a future multichannel decoder should be able to decode the basic 2/0-stereo bitstream properly.

The compatibility is realised by exploiting the ancillary data-field of the MPEG-1 audio frame for the provision of additional channels (see Fig. 5). The variable length of the ancillary data field gives the possibility of carrying the complete multichannel extension information. A standard two-channel MPEG-1 audio decoder just ignores this part of the ancillary data field.

One example of this strategy is the Digital Audio Broadcast system (DAB, developed by the EUREKA 147 Consortium), which will not provide multichannel sound in the first instance. In this case, the multichannel sound system has to be backwards/forwards compatible with an MPEG-1 Audio decoder.

There will be other applications which do not require backwards/forwards compatibility with existing 2/0-stereo sound formats. In these cases, the compatibility

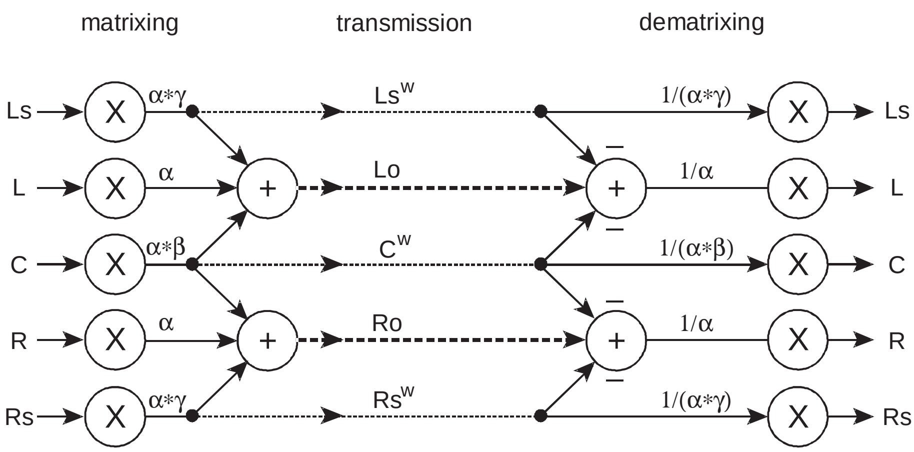

Fig. 4 - The compatibility matrix (encoder) to create the compatible basic stereo signal (Lo, Ro), and the inverse matrix (decoder) to recreate five audio channels.

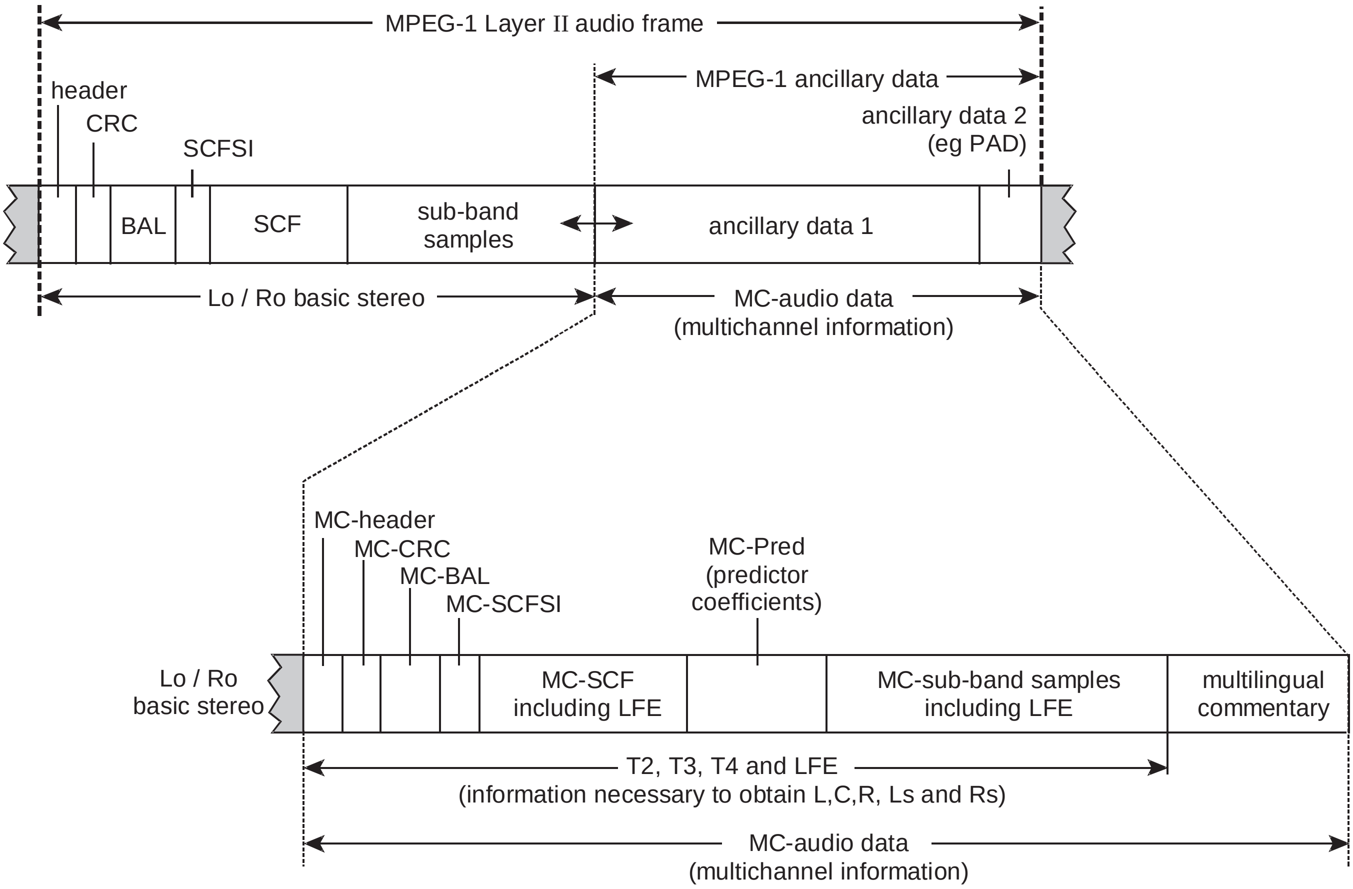

Fig. 5 - Ancillary data field of the MPEG-1 Layer II frame carrying multichannel extension information.

requirement may not be appropriate, because possible coding constraints resulting from the use of compatibility matrixing[18] can be avoided. In order to ensure maximum flexibility and coding efficiency for the different application areas, it seems advantageous to realise both strategies in a universal codec. This is possible by switching the compatibility matrix on or off. In other words, a multichannel sound codec could be used in two different modes. The first being a mode where the basic stereo information consists of a left and right channel that constitute an appropriate downmix of the audio information from all source channels; the second being an optional mode where the basic stereo information may consist only of the left and right channel of the multichannel sound configuration.

The MPEG-2 audio frame may be divided into two parts to accommodate the requirement for the wide range of bit-rates, which extend up to a maximum of 1066 kbit/s (referred to in section 4.1.2). The first part comprises the MPEG-1-compatible part of the bit stream, which provides for Layer I, at a bit-rate of up to 448 kbit/s; for Layer II, at a bit-rate of up to 384 kbit/s; and for Layer III, at a bit-rate of up to 320 kbit/s. The divided frame is shown in Fig. 6 *(overleaf)*. In order to guarantee backwards compatibility, the basic stereo signal must be kept in the MPEG-1 compatible part of the bit stream.

4.1.6 Low frequency enhancement channel

According to ITU-R Recommendation BS. 775,[16] the 3/2-stereo sound format should be able to provide an optional low frequency enhancement (LFE) channel in addition to the full range main channels, being capable of carrying signals in the frequency range 20 Hz to 120 Hz (shown in Fig. 7 *(overleaf)*). The purpose of this channel is to enable listeners who so choose to extend the low frequency content of the programme in terms of both frequency and level. In this way it is the same as the subwoofer channel used in the digital film sound format, and thus reproduction of film sound material would be enhanced in this respect.

4.1.7 Associated services and configurability

For HDTV applications particularly, there is likely to be a requirement for services such as multilingual dialogues, narrative or commentaries associated with the picture, in addition to the main service (see Fig. 8 *(overleaf)*). There are many possibilities; for example, a bilingual 2/0-stereo programme or a 2/0-, 3/0- or 2/1-stereo sound signal plus ‘clean’ dialogue for the hard-of-hearing, a commentary for viewers with poor sight, or for multilingual commentaries.

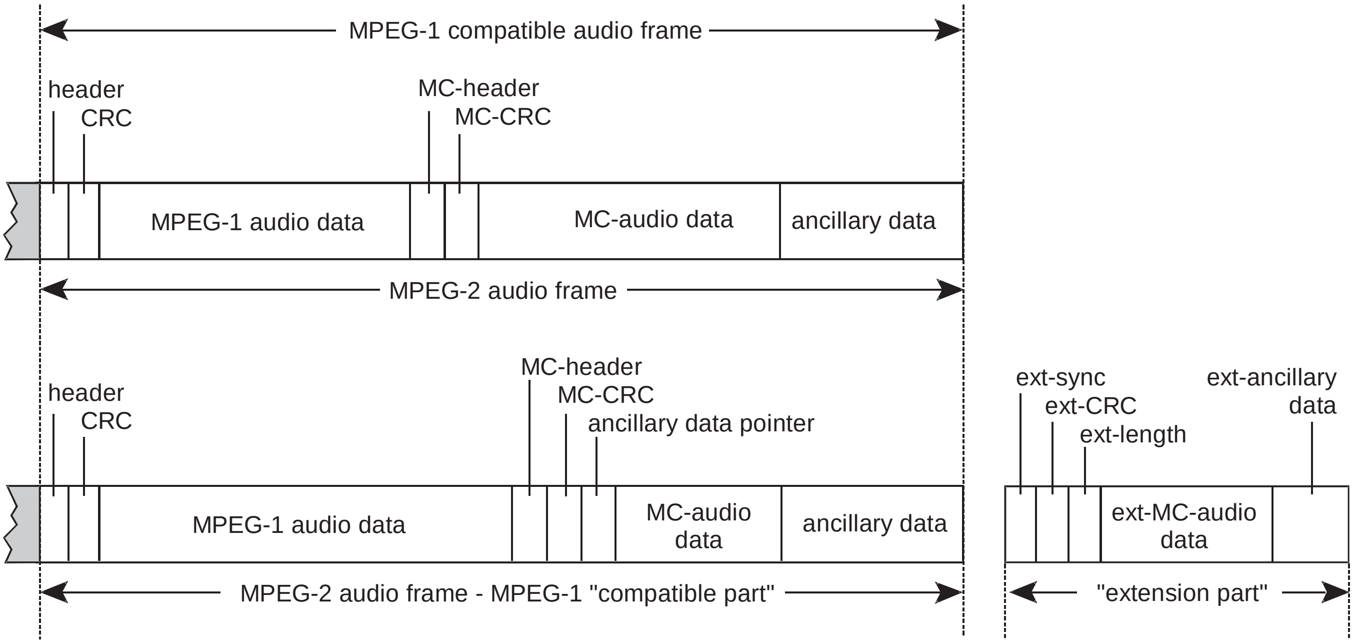

Fig. 6 - MPEG audio frame consisting of the MPEG-1 compatible part and the extension part.

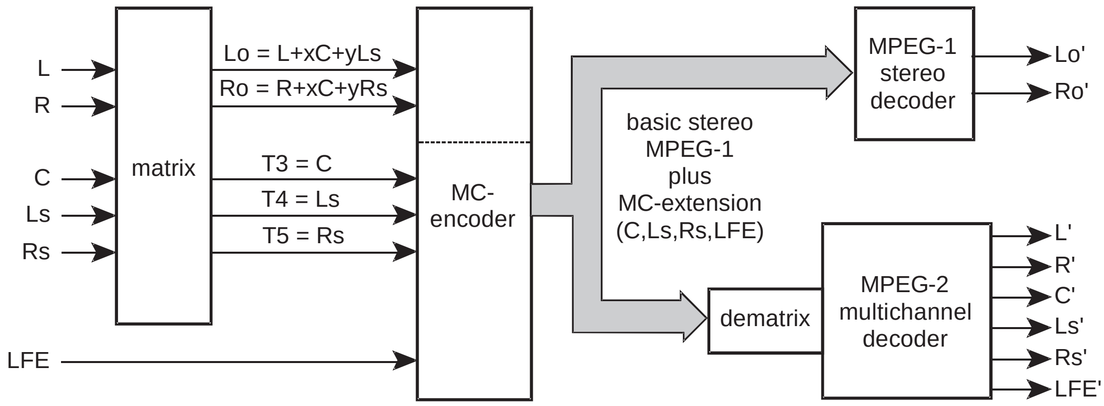

Fig. 7 - Multichannel arrangement incorporating low frequency enhancements (LFE) channel, and the backwards compatibility to MPEG-1

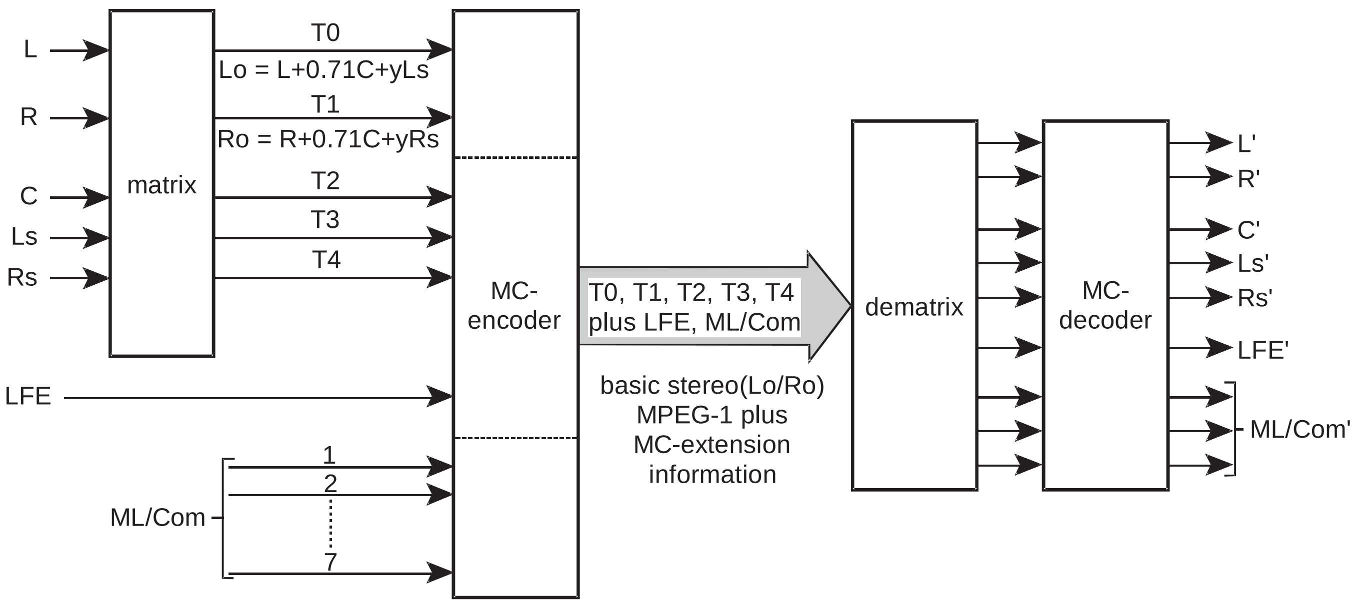

Fig. 8 - MPEG-2 multichannel audio including low frequency enhancement and multilingual channels.

A multilingual service can be provided readily, in combination with surround sound, when the spoken contribution is not part of the acoustic environment that is being portrayed. For example, at a sporting event, surround sound effects plus multiple language mono commentary channels may be provided relatively easily. In contrast, surround sound with drama would require a new multichannel mix for each additional language.

An important issue for multichannel programmes is certainly the 'final mix in the decoder'. This is the reproduction of one selected commentary/dialogue (e.g. via centre loudspeaker) together with the common music/effect stereo downmix (examples are documentary films and sports reportage). If backwards compatibility is required, the basic signals have to contain the information of the primary commentary or dialogue signal, which has to be subtracted in the multichannel decoder when an alternative commentary or dialogue is selected.

In addition to these services, broadcasters should also be considering services for the hard-of-hearing and for viewers with poor sight. In the case of those with hearing difficulties, a clean dialogue channel (i.e. a channel without sound effects) is the most advantageous. For those with poor sight, a descriptive channel is needed. In both cases, these services could be transmitted at a relatively low bit-rate, which would make very little demand on the limited capacity of the transmission channel. Purpose-built receiving equipment would be required for these special services and may not be affordable by the audience to whom it is directed, unless the hardware is economically priced. It may therefore be important to consider a simple solution in the form of a service component that can be decoded by the standard receiving equipment.

Optimum exploitation of the available bit-rate for multichannel stereo performance and sound quality on the one hand, and bilingual programmes or associated services on the other, depends on the application, on the type of programme, etc. For this reason, it is beneficial to have a number of alternative service and quality-level configurations available.

4.2 Composite coding strategies for multichannel audio

If the composite coding methods used for an audio programme deal with more than one channel, the bit-rate required does not necessarily increase proportionally with the number of channels. For multichannel audio, the composite coding technique is very efficient, because there are many correlations, both in the signal itself, and in the binaural perception of such a signal. In the composite coding mode, the irrelevant and redundant portions of the stereophonic signals are eliminated as far as possible. The effects described below may be used to advantage.

4.2.1 Redundancy reduction

Certain stereophonic signals contain interchannel coherent portions, which, in principle, could be transmitted via one channel instead of two.

4.2.2 Common bit pool

The bit-rate per channel required for perceptual coding depends on the signal. It varies dynamically in the region of about 100 kbit/s. Since the individual dynamic bit-rates of the centre and surround signals may not be completely correlated (they may even be uncorrelated), the peaks in the overall bit-rate peaks are generally lower than the sum of the peaks in the individual channels. This means that a common bit pool may be shared between channels ("bit exchange") to give efficient coding.

4.2.3 Dynamic crosstalk

Those parts of the stereophonic signals which are irrelevant with respect to the spatial perception of the stereophonic presentation, are identified by a model of binaural hearing in the encoder. These components are not necessarily masked by the masking characteristics of the ear. But, on the other hand, they do not contribute to the localisation of sound sources; they are ignored in the binaural processes of the auditory system. Therefore, *stereo-irrelevant* components of any stereo signal (L, C, R, Ls or Rs) may be reproduced via *any* loudspeaker, or via several loudspeakers of the arrangement, *without* affecting the stereophonic impression.

4.2.4 Common masking threshold

In the encoder, the individual ('intra-channel') masking thresholds for each of the five input sound signals L/C/R/Ls/Rs are calculated in the same way as in the basic stereo MUSICAM encoder. However, the sub-band samples in the individual channels are quantised according to the highest threshold, taking account of the 'inter-channel' masking effect, called 'Masking Level Difference' (MLD).[19] This is characterised by a decreasing masking threshold when the masker is spatially separated from the source being masked.

The use of the common masking threshold instead of the intra-channel masking threshold, implies that the loudspeaker arrangement and the maximum listening area have to be taken into account. Listening very closely at one loudspeaker may result in the perception of coding noise. Therefore this algorithm is used only as a last resort, when the bit-rate is insufficient. If the peaks of the dynamically-varying bit-rate requirement are higher than the available bit-rate, the optimum

combination of the dynamic crosstalk and the common masking threshold coding method is selected in the encoder. It is suggested that it should be possible to avoid the perception of coding noise, even for extreme locations of the listener, and that small impairments of the stereophonic quality; also, only slightly annoying for some types of programme. For example, the perceived difference between 3/2-stereo and 3/1-stereo presentation of concert-hall music has been found to be very small.[20]

4.3 MPEG-2 extension to lower sampling frequencies

The second objective of MPEG-2 audio was the extension of MPEG-1 coding to include lower sampling frequencies. This extension is particularly useful for applications with bit-rates of 32 to 64 kbit/s per channel and where a bandwidth of 11.25 kHz (with a sampling frequency of 24 kHz) can be accepted. Possible applications include:

- transmission of wideband speech and medium quality audio
- commentary
- distribution of programmes to AM and short wave transmitters
- news channels on the DAB system
- telecommunications.

The improvement in quality results from a better frequency resolution of the polyphase sub-band filter bank in the low and medium frequency region. The quantising distortions can be adapted much more closely to the masking threshold of the audio signals. The improvement for MPEG-2 Layer I and II is greater than for MPEG-2 Layer III, because the MPEG-1 Layer III already has a good frequency resolution in the lower and medium frequency range; improving the frequency resolution by using a lower sampling frequency does not provide such a big advantage.

The coding gain at a sampling frequency of 24 kHz, expressed in kbit/s, can be calculated for the 32 sub-band polyphase filterbank of Layer I and II. The calculation,[21] is based on applying the upper and lower slopes of the masking threshold for the most critical signal, to determine the necessary bit allocation per sub-band. Compared to a sampling frequency of 48 kHz, the coding gain for 24 kHz is about 58 kbit/s per audio channel.

Several international tests have shown that, with 48 kHz sampling frequency, MPEG-1 Audio Layer II, at a bit-rate of 112 to 128 kbit/s per channel, achieves a subjective quality which cannot be distinguished from the original.[22,23] Taking into account the coding gain of the smaller sub-bands for the MPEG-2 Layer II having the half sampling frequency, the full audio quality of a 11.25 kHz low-pass filtered signal can be preserved in a bit-rate of 54 to 70 kbit/s per channel. Tests in MPEG have shown that both MPEG-2 audio Layer II and Layer III meet the requirements for a commentary codec, that is, there are only small audible differences in the subjective quality for coded speech signals, compared to the original with 20 kHz audio bandwidth, even with a bit-rate of 56 kbit/s per audio channel (see Fig. 9).

Compared with MPEG-1, only a few changes have to be made to accommodate reduced sampling frequencies. In the case of Layer I and Layer II (only two decoder tables) the bit-rate and the bit allocation table have to be changed. For these layers, the structure of the bit stream is not changed, and the same type of framing is used. The resulting frame length for 24 kHz sampling frequency is 16 ms for Layer I and 48 ms for Layer II. With Layer III, the number of samples in the frame is halved so that the frame period remains at 24 ms.

The improvement in quality involves no increase in the complexity of the MPEG-1 audio coding for any of the layers. To avoid the need to accommodate additional sampling frequencies in digital audio equipment connected to MPEG-2 codecs, a simple sub-sampling filter needs to be used at the encoder input, together with an up-sampling filter at the decoder output, to maintain 48 kHz sampling at the interconnections.

5. CONCLUSIONS

The concept of perceptual sub-band coding has strongly influenced the MPEG audio group in the standardisation of a generic audio coding scheme. The

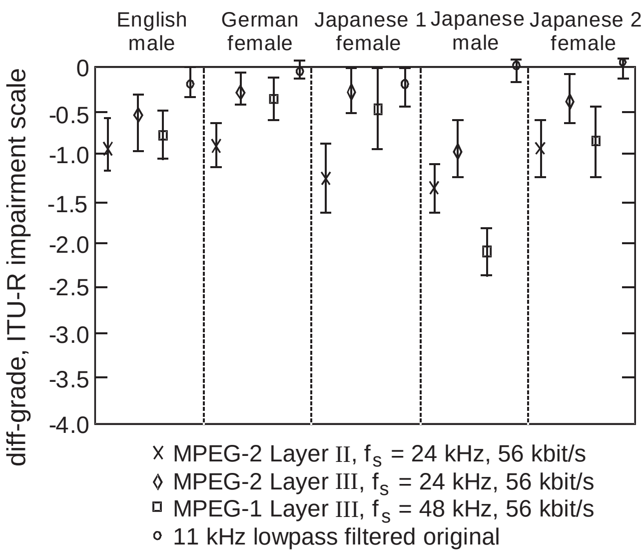

Fig. 9 - Subjective quality, shown in diff-grades for MPEG-2 lower sampling frequency at a bit-rate of 56 kbit/s per audio channel.

aim of MPEG-1 Audio was to establish a coding technique which could be used either with, or quite independently from, the picture coding scheme, with the capability to code high quality audio signals in the range from 192 kbit/s down to about 130 kbit/s per monophonic programme. The higher bit-rates provide some margin for cascading and post-processing. The extension of the work started under MPEG-1, to cover multichannel audio and the use of lower sampling frequencies, forms MPEG-2 Audio.

Digital television will undoubtedly be accompanied by bit-rate-reduced digital audio signals. The choice of coding schemes will have to take account of many factors. Whichever system is chosen, the decisions for a digital transmission format will affect consumers for many decades. Those decisions must therefore be able to stand the test of time. MPEG-1 and MPEG-2 audio, with its generic coding architecture,[24] is certainly a worthy candidate. The first phase of the development of high quality audio coding for widespread use in broadcasting, telecommunication, computer and consumer applications has been completed, with the publication of ISO/IEC 11172-3 (MPEG-1). But the finalisation of MPEG-1 is not the end for standardisation of high quality audio coding systems. MPEG-2 audio multichannel coding, ensuring forwards and backwards compatibility with MPEG-1 mono and stereo encoded audio signals, and MPEG-2 audio low sampling-frequency coding, are designed for an even wider range of applications with and without an accompanying picture.

6. ACKNOWLEDGEMENTS

The authors wish to acknowledge the work of all their colleagues from the Centre Commun d'Etude de Télédiffusion et Télécommunications (CCETT), France; the Fraunhofer Institute for Integrated Circuits, Germany; Philips Consumer Electronics, The Netherlands; Thomson Consumer Electronics, Germany; the University of Hannover, Germany. All of these companies and groups have contributed to the work of the MPEG-2 Audio standard. They also wish to thank Prof. Noll from the Technical University of Berlin for his rôle as the chairman of the MPEG-2 Audio group during the period in which the MPEG-2 standard was defined.

7. REFERENCES

1. ISO/IEC 11172-3, 1992. Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mbit/s – Audio Part. International Standard.

2. ELY, S.R., 1996. MPEG VIDEO CODING: A basic tutorial introduction. BBC Research & Development Report No. 1996/3.

3. SARGINSON, P.A., 1996. MPEG-2: Overview of the systems layer. BBC Research & Development Report No. 1996/12.

4. MEARES, D.J., 1993. Perceptual attributes of multichannel sound. AES 12th International - Conference – The perception of reproduced sound. Copenhagen. Conference proceedings, pp. 171-179.

5. THEILE, G., 1993. Trends and activities in the development of multichannel sound systems. AES 12th International Conference – The perception of reproduced sound. Copenhagen. Conference proceedings, pp. 180-187.

6. ITU RADIOCOMMUNICATIONS SECTOR, 1994. Recommendation BS. 562. Subjective assessment of sound quality. Green Book **1994 BS series volume**, pp. 183-188.

7. TSUTSUI, K., SUZUKI, H., SHIMOYOSHI, O., SONOHARA, M., AKAGIRI, K. *and* HEDDLE, R.M., 1996. ATRAC: Adaptive transform coding for MiniDisc. Audio Engineering Society (AES), special publication, Collected Papers on Digital Audio Bit-rate Reduction.

8. CCIR *(since 1994, the ITU Radiocommunications Sector)*, 1992. Recommendation 712. High-quality sound/data standards for the broadcasting-satellite service in the 12 GHz band. Green Book **RBO Series**, BroadcastingSatellite Service (Sound and Television), Geneva, pp. 20-67.

9. JOHNSTON, J.D. *and* FERREIRA, A.J., 1992. Sum-difference stereo transform coding. Proc. ICASSP, IEEE.

10. INDEPENDENT BROADCASTING AUTHORITY (IBA), BRITISH RADIO EQUIPMENT MANUFACTURERS' ASSOCIATION (BREMA) *and* BRITISH BROADCASTING CORPORATION (BBC), 1988. NICAM 728: Specification for two additional digital sound channels with System I television. ISBN 0 563 20716 7. *First published August 1988.*

11. ISO/IEC DIS 13818-3, 1994. Generic coding of moving pictures and associated audio – audio part. Draft International Standard, March.

12. BOSI, M., 1994. Independent channel and composite audio transform coding: An overview of AC-2 and AC-3. Audio and video digital radio broadcasting systems and techniques. De Gaudenzi, R, *and* Luise, M. (editors), *Elsevier Science* , pp. 73-88.

13. WYLIE, F., 1995. apt X-100: Low-delay, low bit-rate sub-band ADPCM digital audio coding. Audio Engineering Society (AES), special publication, Collected Papers on Digital Audio Bit-rate Reduction.

14. BRANDENBURG, K. *and* STOLL, G., 1996. ISO/MPEG-1 audio: A generic standard for coding of high quality digital audio. Audio Engineering Society (AES), special publication, Collected Papers on Digital Audio Bit-rate Reduction.

15. CHIARIGLIONE, L., et al. 1991. Press release of ISO/MPEG. Document ISO/IEC JTC1/SC29/ WG11 MPEG91/346, November.

16. ITU RADIOCOMMUNICATIONS SECTOR, 1994. Recommendation BS. 775. Multichannel stereophonic sound system with and without accompanying picture. Green Book **1994 BS series**, pp. 257-266.

17. MEARES, D.J., 1991. High-quality sound for high definition television. AES 10th International Conference, Images of Audio, (London, September). Conference digest, pp. 163-177.

18. TEN KATE, W., BOERS, P., MÄKIVIRTA, A., KUUSAMA, J., CHRISTENSEN, K.-E., *and* SØRENSEN, E., 1992. Matrixing of bit-rate - reduced signals. Proc. ICASSP 1992, **2**, pp. 205-208.

19. BLAUERT, J., 1983. Spatial Hearing – The psychophysics of human sound localisation. The MIT-Press, Cambridge, Mass, ISBN 0-262-02190-0.

20. THEILE, G., 1996. HDTV sound systems: How many channels? AES 10th International. Conference, Images of Audio (London, September). Conference digest, pp. 147-162.

21. THEILE G., STOLL G., *and* LINK M.,1988. Low bit-rate coding of high-quality audio signals. An introduction to the MASCAM system, EBU Review Technical No. 230, August, pp. 158-181.

22. ITU RADIOCOMMUNICATIONS SECTOR, 1991. The SR report on the MPEG/Audio subjective listening test (Stockholm April/May). Document JIWP 10-CMTT/1-45.

23. GRUSEC, T., SOULODRE, G. *and* THIBAULT, L., 1995. EIA/NRSC DAR systems subjective tests of audio quality and transmission impairments. Electronic Industries Association, Digital Audio Radio Subcommittee, report on digital audio radio laboratory tests, **1**, published by *EIA Consumer Electronics Group*.

24. STOLL, G., THEILE, G., NIELSEN, S., SILZLE, A., LINK, M., SEDLMEYER, R. *and* BREFORT, A., 1993. Extension of ISO/MPEG-audio layer II to multichannel coding: The Future standard for Broadcasting, Telecommunication and Multimedia Applications. 94th AES Convention (Berlin, March). AES Preprint No. 3550.

Published by BBC RESEARCH & DEVELOPMENT DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP